DOCUMENT RESUME

ED 414 318 TM 027 820

AUTHOR Shrader, Vincent D.

TITLE Designing a Teacher/Course Assessment Instrument for Distance Education.

PUB DATE 1997-10-00

NOTE 9p.; Paper presented at the Annual Meeting of the Northern Rocky Mountain Educational Research Association (Jackson, WY, October 1997).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS College Faculty; *College Students; *Course Evaluation; *Distance Education; *High School Students; High Schools; Higher Education; *Independent Study; *Item Analysis; Measurement Techniques; Student Attitudes; Test Construction

IDENTIFIERS *Brigham Young University UT

ABSTRACT

Independent Study (IS) at Brigham Young University (Utah) enrolls nearly 17,000 high school students and 13,000 college students each year. Results of student evaluations are used to help determine to what extent the IS courses and instructors are meeting student needs. IS administrators decided to redesign the student forms to provide a better measure of course effectiveness and student attitudes toward registration, service, and course materials. A study of about 300 completed forms helped identify trends in the data. Data from this analysis and from student interviews were used to construct new items that were assimilated into a possible item pool. Faculty members were asked to evaluate these items grouped into subscales. Approximately 50 faculty members responded to the request, and about 75 later replied to a request to evaluate a revised instrument. The developed instrument was then sent with each final examination packet until over 500 student forms were returned. An error in the key provided with the form resulted in a second field test with 116 forms for item analysis. This analysis showed that almost all the items had moderate to high adjusted item-total correlation coefficients. A refinement resulted in five subscales, and another item analysis supported these subscales. The instrument was accepted for use in IS evaluations. (Contains eight tables and eight references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 318

Designing a Teacher/Course Assessment Instrument for Distance Education

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
Vincent Shrader

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

by

Vincent E. Shrader
Instructional Psychology and Technology Department
Brigham Young University
Provo, Utah

TM027820

Presented at the Annual Meeting of the
Northern Rocky Mountain Educational Research
Jackson, Wyoming
October 1997

BEST COPY AVAILABLE

Purpose

BYU Independent Study currently enrolls nearly 17,000 high school students and 13,000 college students each year. All students who enroll in Independent Study (IS) courses are asked at the conclusion of the course to evaluate the course and instructor by completing a two-page questionnaire.

The results of these student evaluations help determine to what extent the Independent Study courses and instructors are meeting students' needs. Instructors are regularly given a Course Evaluation Report which lists mean scores for each question on the student evaluation form, overall course and instructor ratings, and student demographics.

The data collected by this process are used to guide the instructional designers at IS in making decisions about course revision and development. The data are also used to help monitor instructors' performance and alert the instructional designers at IS to any problems with the faculty that need to be addressed. Because BYU Independent Study is looked upon as a leader in the field of distance learning, the final form of this measurement instrument should be of great interest to other distance learning institutions throughout the United States.

Justification

The measurement instrument previously used by IS was an adaptation of the student evaluation form collected by the Teacher Evaluation Office on campus and administered to students in each class at the end of the semester. IS retained the format of this form so that the IS form could be read at the Testing Center on campus without modifications and additional expense. Some of the questions were changed and some were deleted to more closely fit the needs of IS.

After five years of using this instrument, the IS administrators decided to redesign the forms. Many questions were simply unsuitable to IS courses and were serving no useful purpose. Many questions were vague or poorly worded, and there were several format problems that reduced the readability of the questionnaire. In addition, the previous form inadequately measured or neglected to measure students' attitudes on areas such as registration, service, and course materials.

Description of the Procedures Used

First it was necessary to refine the constructs which IS, or the client, wished to measure. A content analysis was conducted through a series of interviews with instructional designers, administrators, faculty, and students.

A thorough study of about 300 completed forms of the current evaluation instrument help identify trends in the data. The researcher identified areas where students voiced concerns about the course, instructor, and the service they received from IS staff. The data from this analysis and the interviews were used to construct new items which were assimilated into a possible item pool.

The researcher met with the client for a group interview to gather critical incidents, a list of behaviors which characterize extremes of the performance continuum for the constructs of interest. Questions were asked such as, "Describe a situation where instructor feedback is perfect," and "Describe a situation where instructor feedback is unacceptable."

With the information thus far gathered, 11 separate subscales were developed which covered all the claims, concerns, and issues which had been recorded. All existing items were categorized into those subscales. New items were later developed to place into the subscales so that each subscale contained several items.

The researcher asked the client to rate and rank each subscale in importance. This caused some

consternation among the group. "How can we rank these subscales," they asked, "when all of them are important?" As a result, most subscales were ranked 1's with hardly any 2's or 3's. The ranking also seemed to be difficult for them because they felt that ranking an item last would mean they didn't think the item was important. Conversely, they were concerned that if we used all the items the final form would be too long.

The constructs and the items that represented them were refined again, and a form was sent to each of the IS Faculty asking them to rate and rank the refined list of subscales. They were also requested to define any issues that were not represented. None of the faculty defined any additional issues. Approximately 50 out of 200 forms were returned completed.

The items were refined once more by eliminating some of the lower ranked items in each subscale and revising the wording in some of the existing items. After careful review it was decided to collapse one subscale into a related subscale and totally eliminate another subscale which lacked overall significance leaving nine subscales remaining.

Another meeting was convened with the client. Again, the group was asked to rank the items within each subscale. They were instructed that, although each of the items might be important, they had to be discriminating in order to weed out a number of items for the preliminary tryout. This time the ranking went more smoothly.

The preliminary tryout began a few days later and approximately 75 completed forms were obtained. Exit interviews were conducted with the students as they came out of the testing room. They were asked which questions they felt were vague or unclear and also if there were issues they felt were not represented in the form. Unfortunately, the students didn't provide much helpful feedback. Minor revisions were made in the wording of a few items to prepare the form for an official field test.

The field test was sent out with each final exam packet until over 500 completed forms were returned. While reviewing one of the completed forms, an error was discovered in the key provided on the form. After discussing the problem with a statistician and the client, we decided to readminister the field test.

The second field test yielded slightly more than one fifth the original sample size. This number was disappointing, but time and expense necessitated the immediate development of the new form. With only 116 cases, an item analysis of the data was conducted.

Results

The item analysis from the field test showed that almost all of the items had moderate (.40 to .60) to high (.60 and above) adjusted item-total correlation coefficients. After examining the results and conferring with the client, the researcher collapsed four subscales which seemed to measure the same construct into one subscale. This left five subscales remaining. Another item analysis was performed.

Tables 1-5 describe each item in the final item analysis with its mean, standard deviation, adjusted correlation within its subscale, adjusted correlation within the total scale, and the recommendation to accept or reject the item based on the two adjusted correlation coefficients. The adjusted item-total correlation within its subscale is important if the client desires to have strong correlations within subscales. These items had adjusted item-total correlation coefficients .60 and higher. The adjusted item-total correlation across the total scale is important if the client desires to look at the "whole picture" or how the item relates to all the items in the scale. These items had adjusted item-total correlation coefficients of .50 and higher. These tables have been arranged by subscale.

TABLE 1. Item Statistics For Subscale 1

#	Item	Mean	SD	Adj. r Within Subscale	Adj. r Total Scale	Decision: Accept-A Reject-R
1.	This course met my needs.	4.34	.75	.64	.51	A/A
2.	This course was sufficiently challenging but not too demanding.	3.91	.94	.63	.55	A/A
3.	The educational value of this course was equivalent to a successful classroom experience.	4.03	.95	.75	.66	A/A
4.	In terms of difficulty, this course was just about right (not too easy, not too hard).	4.06	.87	.65	.47	A/R
5.	I learned as much or more in this course than I would have in a similar classroom course.	3.94	1.13	.54	.45	R/R
6.	I would recommend this course to a friend.	4.21	.83	.72	.52	A/A

TABLE 2. Item Statistics For Subscale 2

#	Item	Mean	SD	Adj. r Within Subscale	Adj. r Total Scale	Decision Accept-A Reject-R
7.	The exam questions were clearly worded.	4.19	.82	.62	.35	A/R
8.	The exam(s) accurately represented the course content.	4.09	.87	.71	.51	A/A
9.	The exam(s) concentrated on important points of the course.	4.08	.86	.63	.54	A/A
10.	The exam(s) accurately measured what I learned in the course.	3.89	.98	.73	.42	A/A
11.	The exam questions required me to think rather than just recall trivial information.	3.76	.88	.44	.47	R/R
12.	The exam directions were clear and easy to follow.	4.18	.85	.51	.39	R/R
13.	The "How to Prepare for the Examination" section effectively helped me prepare for the exam(s).	3.84	1.01	.54	.41	R/R

BEST COPY AVAILABLE

TABLE 3. Item Statistics For Subscale 3

#	Item	Mean	SD	Adj. r Within Subscale	Adj. r Total Scale	Decision Accept-A Reject-R
14.	This course was up-to-date.	4.18	.80	.38	.49	R/R
15.	This course was well organized.	4.31	.71	.63	.61	A/A
16.	The course objectives were clearly explained in the course manual.	4.34	.74	.69	.66	A/A
17.	The stated course objectives matched what was actually taught.	4.19	.82	.80	.73	A/A
18.	The page about the instructor helped me develop a sense of enthusiasm for the course.	3.74	.98	.52	.48	R/R
19.	The "How to Succeed" section effectively helped me do well in the course.	3.63	.94	.69	.65	A/A
20.	The lesson self-check questions helped prepare me for assignments and exam(s).	4.25	.83	.40	.42	R/R
21.	The instructions for course assignments were clear and easy to follow.	4.26	.77	.59	.59	R/A
22.	The course materials provided enough information to answer my questions without seeking additional help.	4.22	.81	.49	.49	R/R
23.	My graded assignments were returned promptly.	4.34	.72	.49	.48	R/R
24.	The instructor provided helpful feedback on my assignments.	3.90	.96	.57	.54	R/A
25.	The instructor's comments motivated me to do well in the course.	4.03	.91	.51	.48	R/R
26.	The instructor made positive comments to encourage my progress.	4.26	.82	.67	.59	A/A
27.	The instructor answered my questions clearly.	4.00	.93	.71	.68	A/A
28.	The computer-graded lesson feedback statements were clear and understandable.	3.71	.65	.80	.80	A/A
29.	The computer-graded lesson feedback statements guided me to discover the correct answers to the items I missed.	3.70	.62	.80	.81	A/A
30.	The assignments effectively helped increase my comprehension of the subject matter.	4.23	0.78	0.62	0.67	A/A
31.	The assignments helped me prepare for the exam.	4.29	.79	.63	.65	A/A
32.	The assignments were more than just busywork.	3.90	.88	.71	.68	A/A
33.	The assignments helped me apply what I learned.	4.08	.82	.65	.64	A/A
34.	I received prompt service from the Independent Study staff.	4.24	.80	.67	.63	A/A
35.	The Independent Study staff provided accurate information.	4.38	.65	.59	.56	R/A
36.	The Independent Study staff were courteous.	4.38	.62	.60	.60	A/A
37.	The instructor clearly defined the course requirements at the beginning of the course.	4.32	.73	.72	.68	A/A
38.	The grading procedures for the course seemed fair.	4.38	.63	.71	.66	A/A
39.	The grading criteria were clearly defined in the course manual.	4.32	.73	.61	.59	A/A

BEST COPY AVAILABLE

TABLE 4. Item Statistics For Subscale 4

#	Item	Mean	SD	Adj. r Within Subscale	Adj. r Total Scale	Decision Accept-A Reject-R
40.	The textbook(s) were helpful in learning the material.	4.39	.69	.85	.68	A/A
41.	The textbook(s) were well written and easy to understand.	4.36	.81	.88	.57	A/A
42.	The textbook(s) were up-to-date.	4.26	.88	.74	.51	A/A
43.	The textbook(s) were well suited for independent study.	4.32	.80	.83	.51	A/A

TABLE 5. Item Statistics For Subscale 5

#	Item	Mean	SD	Adj. r Within Subscale	Adj. r Total Scale	Decision Accept-A Reject-R
44.	The exam arrived at the testing site as scheduled.	4.51	.66	.86	.53	A/A
45.	The correct exam was sent to my proctor.	4.59	.56	.91	.54	A/A
46.	The testing facility was free from distractions.	4.17	.97	.51	.42	R/R

Tables 6 and 7 display the intercorrelation between subscales. This is important if the client desires to have subscales which are independent of each other. Table 6 represents the original nine subscale analysis and Table 7 represents the final five subscale analysis. The greater number of high correlations among subscales in Table 6 as compared to Table 7 suggests that the four subscales which were collapsed into one subscale largely measured the same construct. High correlation among subscales is represented by coefficients of .60 and higher.

TABLE 6. Correlations Among Nine Subscales

Subscale	1	2	3	4	5	6	7	8
2	.41							
3	.64	.55						
4	.47	.56	.65					
5	.55	.47	.68	.61				
6	.50	.22	.58	.45	.54			
7	.32	.49	.70	.58	.64	.61		
8	.40	.32	.42	.42	.62	.49	.62	
9	.59	.21	.46	.35	.49	.47	.35	.40

BEST COPY AVAILABLE

TABLE 7. Correlations Among Five Subscales

Subscale	1	2	3	4
2	.41			
3	.62	.58		
4	.40	.32	.58	
5	.59	.21	.50	.40

Table 8 displays each subscale with its corresponding construct, number of items, coefficient alpha, mean score, and standard deviation.

TABLE 8. Subscale Descriptions and Reliability Coefficients

Sub-scale	Construct	Number of Items	Alpha Reliability Coefficient	Mean Score	Standard Deviation
1	Over-all experience is positive and productive.	6	0.81	24.48	3.95
2	Exams effectively measure knowledge.	7	0.79	28.03	4.18
3	Course materials, assignments, and feedback contribute to a successful learning experience.	26	0.94	107.57	13.07
4	Textbooks are an integral part of class.	4	0.89	17.32	2.77
5	Testing process is efficient.	3	0.95	13.27	1.82

Validity of the instrument was limited to content validity using the director and the instructional designers at BYU Independent Study as content experts. They were responsible for judging whether the items in the instrument adequately sampled the domains of interest.

Conclusions and Recommendations

This project is an affirmation that real life projects do not always adhere strictly to the theoretical guidelines taught in textbooks. In real life applications there may be a smaller sample size than is preferred, or the client might not always accept what the analysis suggests. In this case, the options presented to the client based on the analysis seemed to confirm pre-existing ideas about which were good or bad items.

The client decided to keep a few of the "bad" items despite the item analysis simply because he considered it important to know the responses to those items. The client was made aware that keeping these items could sacrifice the reliability of the instrument. He said he understood the implications. After all the revisions had been made, the instrument was trimmed down to 26 items. The client was extremely pleased with the finished draft. The client was offered the following recommendations:

1. The client should be mindful of those items that seem dubious so that revisions can be made in the future if necessary.
2. An item analysis with a larger sample should be conducted at a later time when more students are enrolled. The new analysis should be compared with the previous analysis to see if the results are similar.
3. New items should be tried out and considered for inclusion in the instrument as the need arises.

BEST COPY AVAILABLE

References

Arreola, R. A. & Lawrence, M.A. (1991). Developing a Comprehensive Faculty Evaluation System. Tucson: Comprehensive Data Evaluation Services, Inc.

Centra, J. A. (1979). Determining Faculty Effectiveness : Assessing Teaching,Research, and Service for Personnel Decisions and Improvement. San Francisco: Jossey-Bass.

Centra, J. A. (1993). Reflective Faculty Evaluation : Enhancing Teaching and Determining Faculty Effectiveness. San Francisco: Jossey-Bass.

Crocker, L. & Algina, J. (1986). Introduction to Classical and Modern Test Theory. Fort Worth: Holt, Rinehart and Winston.

Doyle, K. O. (1983). Evaluating Teaching. Lexington, Mass: Lexington Books.

Miller, R. I. (1972). Evaluating Faculty Performance. San Francisco: Jossey-Bass.

Spector, P. E. (1992). Summated Rating Scale Construction: An Introduction. Newbury Park: Sage Publications Inc.

Worthen, B. W., Borg, W. R. & White, K. R. (1993). Measurement and Evaluation in the Schools. New York: Longman.